DOCUMENT RESUME

ED 202 876 TM 810 297

AUTHOR Marsh, David D.; And Others
TITLE Comparing Ideal and Real: A Study of Evaluation Practice in California Using the Joint Committee's Evaluation Standards.
PUB DATE Apr 81
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; *Evaluation Criteria; Evaluators; *Program Evaluation; *Program Implementation; *Standards
IDENTIFIERS *California; *Evaluation Problems

ABSTRACT

The Joint Committee on Standards for Educational Evaluation, under the direction of Daniel Stuffelbeam, completed standards for the practice of evaluation. Little is known about the extent of congruence between evaluation practice and the ideal as represented by the standards. The purpose of the study was to obtain information regarding how evaluation practitioners who are members of the California Association of Program Evaluators throughout the State, perceive the: (1) extent to which each standard is an ideal to be sought; (2) degree to which each standard is currently being performed by practitioners; and (3) major reasons why selected standards are currently not being met. An overall average of 97.6% of the respondents indicated the thirty standards as a whole should be retained as important ideals to guide the process and practice of evaluating school programs. The overall perception of the degree to which each standard is currently being performed was indicated to be less than a 3.0 rating on a five-point scale. The predominate single reason provided for low quality practice was a combination of evaluator incompetence, especially regarding standards emphasizing technical or design features, and bureaucratic/political constraints. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED202876

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

COMPARING IDEAL AND REAL:

A STUDY OF EVALUATION PRACTICE IN CALIFORNIA

USING THE JOINT COMMITTEE'S EVALUATION STANDARDS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
D Marsh
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

David D. Marsh
University of Southern California

Warren B. Newman
Office of Los Angeles County
Superintendent of Schools

Bill F. Boyer
University of Southern California

TM 810 397

A paper presented at the Annual Conference of the American Educational Research Association in Los Angeles, April 13-17, 1981.

This research utilized the Standards For Evaluations of Educational Programs, Projects and Materials. The reader is referred to full text of the standards as they appear in Joint Committee on Standards for Educational Evaluation, Standards for Evaluations of Educational Programs, Projects and Materials. New York: McGraw-Hill, 1981.

Introduction: The Joint Committee on Standards for Educational Evaluation, under the direction of Daniel Stufflebeam, recently completed standards for the practice of evaluation (Joint Committee on Standards for Educational Evaluation, 1981). This prestigious committee included representatives from American Educational Research Association, American Psychological Association, National Council of Measurement and Evaluation, as well as practitioner groups such as American Association of School Administrators, Association of Supervisors and Curriculum Developers, National Education Association, American Federation of Teachers, and the Education Commission of the States. Previous annual meetings of AERA provided a forum wherein AERA members could give feedback regarding the wording of the standards and their possible impact upon evaluation theory, policy and practice.

Over the last two years, the standards were field-tested at many sites so that they could receive the widest possible review by a variety of role groups. Important and substantial revisions were made by the Joint Committee based on this review process. For example the statements were revised to cover a wider range of evaluation approaches. This shift is in keeping with the stated focus of the standards reported as, "The Joint Committee attempted to recognize in the standards all types of studies used to evaluate educational materials, programs and projects" (Joint Committee, p. 7).

The standard can be helpful in many ways:

> The Joint Committee foresaw several benefits from the development of sound standards: a common language to facilitate communication and collaboration in evaluation; a set of general rules for dealing with a variety of specific evaluation problems; a conceptual framework by which to study the often confusing world of evaluation; a set of working definitions to guide research and development on the evaluation process; a public statement of the state of the art in educational evaluation; a basis for self-regulation and accountability by professional evaluators; and an aid to developing public credibility for the educational evaluation field. (Joint Committee, p. 5)

The potential benefits represent major contributions to the field of educational evaluation.

Currently, little is known about the extent of congruence between evaluation practice and the ideal as represented by the standards. It also is not known if lack of congruence between practice and ideal is a function of lack of evaluator competence, bureaucratic constraints or other factors. Effective use of the standards as a guide for designing preservice and inservice training for evaluators, and in other applications of the standards, will require careful study of current evaluation practice vis-à-vis the standards.

Objectives. The purpose of the study is to obtain information regarding how evaluation practitioners:

1. Perceive the extent to which each standard is an ideal to be sought.
2. Perceive the degree to which each standard is currently being performed by practitioners.
3. Perceive the major reasons why selected standards are currently not being met.

Given the wide range of topics covered by the standards and the difficulty of obtaining objective information about current practice, the perceptions of experienced practitioners about themselves and their peers provided the best available information about current practice vis-à-vis the standards.

Sample. The population for the study is the membership of the California Association of Program Evaluators (CAPE). The membership consists of school district, county office, state department of education, university faculty, and privately-based evaluators of local education programs in California. Currently, the members represent a sizeable proportion of active educational evaluators in the state. They also represent a group that is very knowledgeable about evaluation practice in California.

All CAPE members were asked to complete the mailed questionnaire. A follow-up letter was used to ensure a high response rate. Questionnaires were mailed to 132 persons resulting in the return of 78 questionnaires. This constitutes a response rate of 59.1%.

Instrumentation. The Joint Committee report describes 30 separate standards. The standards are presented, ". . . in four groups that correspond to four main concerns about any evaluation--its utility, feasibility, propriety and accuracy" (Joint Committee, p. 1). The Joint Committee explains the focus of each cluster of standards as follows:

- The Utility Standards are intended to ensure that an evaluation will serve the practical information needs of given audiences (p. 19)

- The Feasibility Standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal (p. 51)

- The Propriety Standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results (p. 63)

- The Accuracy Standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features of the object being studied that determine its worth or merit (p. 97)

Each standard is explained and clarified through a commentary which includes an overview of intent, guidelines for application, common pitfalls, caveats (or warnings against being overzealous in implementing the standard), and an illustration of the standard's application.

The questionnaire presents only the statement of the standard itself. This limitation may have hindered the full understanding of the respondent about nuances associated with each standard but this limitation was necessary if questionnaire length and respondent burden were to be kept within reasonable bounds. The full report of the standards covers 140 pages. The full list of standards appears in Table 2 in the Appendix A.

The questionnaire lists the evaluation standards within the four categories described above. Respondents were asked to rate each standard according to the extent to which it is sought

as an ideal by the practitioner and the level of quality of current practice in evaluating educational programs. For each standard where current practice were rated less than good by the practitioner, he/she was asked to indicate the major reasons for inadequacy of current performance. The instrument underwent several rounds of pilot testing and review prior to being sent to the CAPE membership. Minor modifications in response categories were made on the basis of the pilot testing. Portions of the questionnaire appear as Appendix B.

Findings. The sample (N=78) of CAPE members who responded to the questionnaire appears to be representative of the overall CAPE membership regarding their "Role" (primary employer) and "Function" (primary job-position). However, compared to all CAPE members, percentages of respondents associated with the last two "Role" categories in Appendix A, Table 1, "Higher Education: 10.3%" and "Private Firm: 20.5%," would have to be reversed in order for the sample to be more representative of the population.

The overall response related to the first question--"How evaluation practitioners perceive the extent to which each standard is an ideal to be sought"--was highly positive regarding the appropriateness, importance and need to retain the thirty standards (See Appendix A, Table 2). An average of 92.1% of the respondents indicated that the "Standards" are "appropriate, important, and should be retained--as stated." And an additional

5.5% indicated that the "Standards: should be retained as important "ideals to be sought" with some revision in the rhetoric associated with designated statements. Therefore, an average of 97.6% of the respondents indicated that the "Standards" as a whole should be retained as important ideals to guide the process and practice of evaluating school programs. Only one of the thirty "Standards" (B2 - "Political Viability"), was recommended for exclusion as an ideal by more than ten percent of the respondents (i.e., 10.8%). Two respondents described their reservations about the political viability standard as follows:

> "I have worked a compromise. I write what I find to be correct and appropriate. The administrators control the distribution of my findings. Administrators have to deal with the 'political climate.' Evaluators should not tailor findings to suit the [political] climate." -- A Public School Practicing Evaluator

> "This statement is too political. The evaluator serves a constituency--a consumer of some service partly for/by the public. The evaluation should focus on the consumer's benefits and needs. If you start in on all the factions, there's no end to lost time, and the reason for the [evaluation] program gets lost." -- A Private Firm Practicing Evaluator

These concerns are typical of respondents' comments about the problems of the political viability standard as an ideal.

The overall response related to the second question--"How evaluation practitioners perceive the degree to which each standard is currently being performed by practitioners"--was considerably less positive. Respondents were asked to assess the quality of current practice vis-à-vis each of the "Standards" in evaluating school programs. Their ratings were

done on a 5-point scale: 1=Poor; 2=Fair; 3=Good; 4=Very Good; and 5=Excellent. As reflected in Appendix A, Table 3, the overall mean score rating of current practice of the thirty Standards was indicated to be less than "Good" (3.0). The average rating was 2.82, with an average of 37% of the respondents indicating the quality of current practice to be "Fair" or "Poor." In short, these respondents perceived current practice as being of only fair to good quality when examined in light of the evaluation standards.

Considering the standards grouped into their four respective categories (Utility, Feasibility, Propriety, and Accuracy), current practice was perceived by respondents as follows:

a.	Propriety:	Mean Score	3.16
		Percentage of "Fair-Poor" Ratings	22.8%
b.	Accuracy:	Mean Score	2.88
		Percentage of "Fair-Poor" Ratings	32.5%
c.	Feasibility:	Mean Score	2.60
		Percentage of "Fair-Poor" Ratings	44.2%
d.	Utility:	Mean Score	2.49
		Percentage of "Fair-Poor" Ratings	54.6%

As reflected in Appendix A, Table 4, the assessment of quality of current practice of the respective Standards range between a lower mean score rating of 2.09 and assessed as "Fair-Poor" by 74.6% of the respondents ("Evaluation Impact"), to as

high as 3.64 and assessed as "Fair-Poor" by only 4.3% of the respondents ("Rights of Human Subjects"). The quality of current practice for various of the standards will be discussed in more detail later in the paper.

Where respondents rated current practice as fair or poor, they were asked to indicate their perceptions as to why practice was of low quality. The questionnaire provided the following response categories:

- BUREAUCRATIC AND PROCEDURAL CONSTRAINTS
- TIME CONSTRAINTS
- BUDGET CONSTRAINTS
- PROGRAM PERSONNEL RESISTANCE
- LEVEL OF EVALUATOR COMPETENCE
- OTHER--(Please specify)

Respondents could check as many categories as applied. The results are presented in Table 3, Appendix A.

The respondents reported that, in general, the predominant reasons for low quality practice in evaluation were:

1. "Level of Evaluator Competence" In 86.7% of the incidents where respondents rated practice as fair or poor, this reason was given as at least one of the main reasons for low quality practice.
2. "Time Constraints" In 70.0% of the incidents where respondents rated practice as fair or poor, this reason was given as at least one of the main reasons for low quality practice.

3. "Bureaucratic Procedural Constraints" In 60% of the incidents where respondents rated practice as fair or poor, this reason was given as at least one of the main reasons for low quality practice.

4. "Program Personnel Resistence" In 60% of the incidents where respondents rated practice as fair or poor, this reason was given as at least one of the main reasons for low quality practice.

These findings reflect the overall belief that respondents attribute low quality practice to a combination of evaluator imcompetence and bureaucratic/political constraints.

It is now possible to provide a more detailed analysis of the perceived quality of current practice linked with perceived reasons for low quality practice which was reported for some of the standards. Respondent assessment of current practice tends to reflect four distinct quality levels of practice (Appendix A, Table 4) which we have labeled as:

a. Satisfactory Practice: receiving a mean score range of 3.64-2.98, and assessed as "Fair-Poor" by 4.3%-20% of the respondents (seven "Standards").

b. Fair Practice: receiving a mean score range of 2.97-2.82, and assessed as "Fair-Poor" by 29.5%-33.8% of the respondents (seven "Standards").

c. Marginal Practice: receiving a mean score range of 2.79-2.63, and assessed as "Fair-Poor" by 39.7%-48.7% of the respondents (ten "Standards").

d. Poor Practice: receiving a mean score range of 2.69-2.09, and assessed as "Fair-Poor" by 50.0%-74.6% of the respondents (six "Standards").

Standards which fall into each of these quality levels of practice, and an analysis of perceived reasons for low quality performance within each level, are as follows:

a. Statisfactory Practice (of 7 "Standards"): an average of 14.0% rated the practices at "Fair-Poor" ($\bar{X}$=3.29).

There were four "Propriety" and three "Accuracy Standards" reflected in this range of practice. They focused primarily on the non-technical and non-controversial political features or concerns of the evaluation process. These practices were related to the more generalized concerns for conducting "legal, ethical, and responsible" evaluations which are "accountable to the general welfare, safety, and rights" of both the public or society, as well as the "worth, dignity and rights of individual persons associated with an evaluation." In this regard, it was perceived that the "purposes, objects, and procedures" of the evaluation process are sufficiently identified and described, and based on "adequate and defensible sources of information" to ascertain the general features and value-merit of the object being studied and evaluated.

b. Fair Practice (of 7 "Standards"): an average of 31.1% rated the practice to "Fair-Poor" ($\bar{X}$=2.92).

Again, the "Propriety" and "Accuracy Standards" (three from each category) prevailed in this range of practice; alongwith one "Feasibility Standard" focused on the employment of "practical evaluation procedures aimed at ensuring minimum disruption" of the educational program being studied (a "prudent and diplomatic" practice). The assessed level or range of practice/application of these "Standards" appears to be rather tightly grouped in that almost one-third (31.0%) of the respondents rated the current practice of each to be "Fair-Poor." While a majority of the respondents perceived these practices to be satisfactorily functioning in the process of evaluating school programs, there appears to be a perceived need for considerable improvement in the practices of these respective "Standards." Likewise as indicated by the nature of the "Standards" reflected in the first category of "Satisfactory Practice," these practices seem to generally focus on the relatively non-technical aspects of collecting, analyzing, and reporting evaluative information. That is, regarding the technical adequacy and accuracy of obtained information, the orientation of these practices focused on the more technical aspects of: 1) ensuing valid and reliable

measurement; and 2) providing "objective (unbiased) and explicitly reported descriptions of what, why, when, and in what context an object was evaluated; what information was obtained from the study; and what justified conclusions were drawn from the findings to guide specific recommendations for decision-making (i.e., "(A4) - Report Clarity").

c. Marginal Practice (of 10-Standards): an Ave. of 43.2% rated the practice at "Fair-Poor" ($\bar{X}$=2.70).

The "Accuracy" and "Utility" Standards which are oriented to ensuring the technical adequacy, accuracy, objectiveness, and credibility of evaluation purposes, procedures, obtained information, and derived/reported conclusions and recommendations, are reflected in this quality level of practice. The perceptions of educational program evaluators regarding these critical issues of the evaluation process reflect a general lack of confidence regarding the technical proficiency level of educational program evaluation. Respondents see incompetence or impotence among many evaluators. It seems that most of the practices inherent in these ten "Standards" are in the closest proximity to direct control by the individual program evaluator. This finding is further substantiated in that the prevailing reason given for the "Fair-Poor" practice/application of these particular Standards (with the exception of: "(C1) Formal Obligation"; and "(A6) Report Discrimination"), was indicated as--"Level of Evaluator Competence."

d. Poor Practice (of 6 "Standards"): an Ave. of 60.0% rated the practice to be "Fair-Good" ($\bar{X}$=2.36).

With the exception of "A3 - Information Scope and Selection" and "(A4) - Valuation Interpretation" (and to some degree, "(A1) - Audience Identification"), it appears that the major reasons for inadequate practice of these Standards are perceived to be outside of the realm of direct control by evaluators. Instead, the reasons given for these inadequacies relate primarily to "Bureaucratic Procedural and Time Constraints" (as well as "Program Personnel Resistence" regarding the practice of "(A8) - Evaluation Impact"). In contrast, the "(A3) and (A4) Standards" are perceived to be in close proximity to the evaluator's direct control; and both are related to the previous group of the more technically oriented practices (ref. D.#.c., above).

It seems that the predominant single reason given for either marginal or poor quality of practice among the respective standards is clearly perceived to be--"Level of Evaluator Competence" (ref. Appendix A, Table 3).

Finally, respondents' perception of the quality of current practice were analyzed by the role and function of the respondent. The results are presented in Table 5. There were no significant differences between respondents who served as practicing evaluators, administrators of evaluation units, program managers, or others in terms of their perception of the quality of current evaluation efforts.

In contrast, respondents differentiated by institutional affiliation frequently differed in their perceptions of the quality of current practice in evaluation. These differences in perception were significant at the .05 level for 18 of the 30 standards. Generally, the most positive ratings were indicated by respondents affiliated with a public school followed in order by higher education personnel, and government agency personnel including county office and state department staff. The most negative ratings stemmed from those affiliated with a private firm.

Discussion. It is clear that respondents saw the standards as representing their ideals in evaluation. It is not apparent whether the respondents would have added other standards but the large majority of respondents agreed with the

standards as presented. The Joint Committee has done a fine job in describing ideals which communicate to practitioners and are affirmed by them.

It is equally clear that respondents perceive practice in evaluation to be of margianl quality vis-à-vis the standards. Such perceptions were held by respondents from different types of institutions and were critical of evaluation practice related to most of the standards. Respondents attributed low quality practice to a combination of evaluator incompetence, especially regarding standards emphasizing technical or design features, and to bureaucratic/political constraints.

It is understandable that practice would be perceived as of low quality at this stage in the evolution of evaluation as an area of professional practice. The field is mature enough that initial hopes for instant success have been dashed while training programs for comsumers and practitioners are not fully developed.

The improvement of evaluation practice may be difficult to achieve. While the experienced practitioners or wise theoretician can describe more realistic expectations for evaluation practice, it is not clear that such realism has been incorporated into the design and organization mileau for new evaluation efforts. Moreover, cuts in federal and state funds may decrease the chances for training and technical assistance programs to continue and improve. Finally, funding cuts may also quicken

the already frantic pace at which most evaluations now are conducted resulting in less time for both technical and political/bureaucratic needs in evaluation.

Implications. Several implications follow from this discussion. The Joint Committee should propose strategies for creating a national network of training and technical assistance efforts aimed at both consumers and conductors of evaluations. Such training will need to cover far more than technical issues in evaluation; they must address the complex political/bureaucratic issues as well. Development of training materials to meet the latter training needs remains an unfinished task. In fact, little research or scholarship is available on these topics.

Funding agencies, whether schools, school districts, or states, should also take note of these findings. Realistic use of the standards as a framework for reviewing evaluation proposals will be needed. Funding agencies must also examine their responsibilities for improving evaluation practice through the establishment of more realistic time schedules, better articulation of decision and audience priorities and closer cooperation with evaluators in the utilization of findings.

Evaluators themselves need additional training as do program leaders and decision-makers. This training ought to focus on the critical actors in an evaluation as a group when the training addresses political or bureaucratic issues.

Given the wide diversity of approaches within evaluation, it probably is not desirable to narrowly describe the content of such training. Yet training and technical assistance obviously are needed if evaluation practice is to improve.

Appendix A

Statistical Tables

Table 1

Description of Respondents by Role and Function

(N=78) FUNCTION	ROLE Public School	Govern-ment	Higher Education	Private Firm	Raw Total
Practicing Evaluator	15 19.2%	8 10.3%	3 9.0%	7 3.8%	33 42.3%
Administrator of Eval. Unit	7 9.0%	3 3.8%	1 1.3%	1 1.3%	12 15.4%
Program Manager	5 6.4%	2 2.6%	3 3.8%	0 0.0%	10 12.8%
Two or More Functions	9 11.5%	5 6.4%	1 1.3%	8 10.3%	23 29.5%
Column N= Total %=	36 46.2%	18 23.1%	8 10.3%	16 20.5%	78 100.0%

Table 2

Respondents' Perception of the Degree to which Each Standard Should be Included As Is, Revised, or Excluded As An Ideal

N = 78 Category of Standard		Statement of Standard	Percentage Indicating Retention of Standard "As Is"	"As Is"/Revised
UTILITY:		The Utility Standards are intended to ensure that an evaluation will serve the pratical information needs of given audiences. These standards are:		
(A1)	*Audience Identification:*	Audiences involved in or affected by the evaluation should be identified, so that their needs can be served.	92.1	98.7
(A2)	*Evaluator Credibility:*	The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance.	92.1	100.0
(A3)	*Information Scope and Selection:*	Information collected should be of such scope and selected in such ways that pertinent questions about the object of the evaluation are addressed, and so that the information is responsive to the needs and interests of specific audiences.	94.7	100.0
(A4)	*Valuational Interpretation:*	The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the base value judgements are clear.	96.0	98.7
(A5)	*Report Clarity:*	The evaluation report should clearly describe the object being evaluated and its context and the purposes, procedures, and findings, so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.	94.9	100.0
(A6)	*Report Dissemination:*	Evaluation findings should be disseminated to clients and other ritht-to-know audiences, so that they can assess and use the findings.	94.7	98.8
(A7)	*Report Timeliness:*	Release of reports should be timely so that audiences can best use the reported information.	91.0	96.4
(A8)	*Evaluation Impact:*	Evaluations should be planned and conducted in ways that encourage follow-through by members of the audience.	87.2	97.5

Category	Standard		
FEASIBILITY:	The Feasibility Standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal. These standards are:		
(B1) *Practical Procedures:*	The evaluation procedures should be practical, so as to ensure that associated disruption is kept to a minimum, and that findings can be obtained.	94.6	98.7
(B2) *Political Viability:*	Evaluations should be planned and conducted with anticipation of the different positions of various interest groups so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.	83.8	89.2
(B3) *Cost Effectiveness:*	The evaluation should produce information of sufficient value to justify the resources used.	82.4	93.2
PROPRIETY:	The Propriety Standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results. These standards are:		
(C1) *Formal Obligation:*	Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renogotiate it.	83.8	91.9
(C2) *Conflict of Interest:*	Conflict of interest should be avoided in evaluation, or should it occur, dealt with openly and honestly, so that it does not compromise the evaluation processes and results.	94.7	97.4
(C3) *Full and Frank Disclosure:*	The evaluator's oral and written reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the evaluation.	97.4	100.0
(C4) *Public's Right to Know:*	The formal parties to an evaluation are responsible to uphold the principles and statutes, such as those dealing with public safety and the right to privacy.	88.3	97.4
(C5) *Rights of Human Subjects:*	Evaluators must design and conduct their evaluations in such ways that the rights and welfare of the human subjects are respected and protected.	94.9	98.7

Category		Standard		
PROPRIETY: (cont)				
(C6)	*Human Interactions:*	Evaluators should be respectful of human dignity and worth in their interactions with other persons associated with an evaluation.	89.7%	92.3%
(C7)	*Balanced Reporting:*	The evaluation should be equitable in its presentation of strengths and weaknesses of the object under investigation, so that strengths can be built upon and problem areas addressed.	92.2	100.0
(C8)	*Fiscal Responsibility:*	The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and should be otherwise prudent and ethically responsible.	93.5	97.4
ACCURACY:		The Accuracy Standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features of the object being studied that determine its merit. These standards are:		
(D1)	*Object Identification:*	The object of the evaluation (program, project, material) should be sufficiently examined, so that the form(s) of the object being considered in the evaluation can be clearly identified.	86.7	96.0
(D2)	*Context Analysis:*	The context in which the program, project, or material exists should be described in enough detail, so that the likely influences of the context on the object will be identified.	90.9	98.7
(D3)	*Described Purposes and Procedures:*	The purposes and procedures of the evaluation should be described in enough detail so that they can be identified and assessed.	96.1	100.0
(D4)	*Defensible Information Sources:*	The sources of information should be described in enough detail, so that adequacy of the information can be assessed.	97.4	100.0
(D5)	*Valid Measurement:*	The information-gathering instruments and procedures should be chosen or developed and implemented in ways that will assure that the interpretation arrived at is valid for the given use.	96.1	100.00

Category		Standard		
ACCURACY: (cont.)				
(D6)	*Reliable Measurement:*	The information-gathering instruments and procedures should be chosen or developed and implemented in ways that will assure that the information obtained is reliable.	93.6%	98.7%
(D7)	*Systematic Data Control:*	The data used in an evaluation should be reviewed and corrected so that evaluation reports will not be needlessly flawed.	97.4	98.7
(D8)	*Analysis of Quantitative Information:*	An evaluation's quantitative information should be appropriately and systematically analyzed to ensure supportable interpretations.	97.4	98.7
(D9)	*Analysis of Qualitative Information:*	An evaluation's qualitative information should be appropriately and systematically analyzed to ensure supportable interpretations.	97.4	98.7
(D10)	*Justified Conclusions:*	The conclusions reached in an evaluation should be explicitly justified, so that the audiences can assess them.	88.5	98.7
(D11)	*Objective Reporting:*	Evaluators should be independent of what is evaluated, and the evaluation procedures should provide safeguards, so that the evaluation findings and reports are not distorted by the personal feelings and biases of any part to the evaluation.	84.4	94.8

Table 3

Respondents' Perception of the Quality of Current Practice on the Standards and Reasons for Low Quality of Performance

N. = 78	"5-Point Scale" Mean Score Rating of Practice	Adjusted Percentage Who Rated Practice as "Fair-Poor"	Freq. of Reasons Given for "Fair-Poor"						
			Burea	Time	Budg	Resist	Comp	Other	
UTILITY:									
(A1) Audience Identification:	2.67	50.0%	18	17	12	14	15	4	
(A2) Evaluator Credibility:	2.73	46.6%	9	9	12	10	27	2	
(A3) Information Scope and Selection:	2.43	55.6%	19	16	15	18	21	2	
(A4) Valuational Interpretation:	2.39	57.1%	17	18	7	15	20	4	
(A5) Report Clarity:	2.70	42.5%	11	16	13	7	20	4	
(A6) Report Dissemination:	2.63	47.1%	24	15	13	16	7	2	
(A7) Report Timeliness:	2.30	63.4%	31	32	14	9	8	2	
(A8) Evaluation Impact:	2.09	74.6%	29	24	15	24	16	6	
FEASIBILITY:									
(B1) Practical Procedures:	2.82	30.3%	11	8	5	6	13	3	
(B2) Political Viability:	2.28	60.7%	25	13	7	20	14	3	
(B3) Cost Effectiveness:	2.70	41.6%	14	10	10	8	10	1	

"5-Point Scale"

Poor	Fair	Good	Very Good	Excel.
1	2	3	4	5

Table 3.(cont.)

N = 78	"5-Point Scale" Mean Score Rating of Practice	Adjusted Percentage Who Rated Practice as "Fair-Poor"	Freq. of Reasons Given for "Fair-Poor"						
			1	2	3	4	5	6	
PROPRIETY:									
(C1) Formal Obligation:	2.66	41.0%	13	12	6	15	8	2	
(C2) Conflict of Interest:	2.97	29.5%	6	2	1	10	8	3	
(C3) Full and Frank Disclosure:	2.97	31.0%	16	6	1	14	7	4	
(C4) Public's Right to Know:	3.45	10.1%	2	1	1	4	3	2	
(C5) Rights of Human Subjects:	3.64	4.3%	0	1	1	0	1	1	
(C6) Human Interactions:	3.44	14.7%	4	3	2	3	8	1	
(C7) Balanced Reporting:	2.94	31.9%	10	7	5	15	11	1	
(C8) Fiscal Responsibility:	3.15	20.0%	7	4	3	4	8	1	
ACCURACY:									
(D1) Object Identification:	2.98	19.1%	4	4	4	5	5	2	
(D2) Context Analysis:	2.71	39.7%	7	11	8	12	14	3	
(D3) Described Purposes and Procedures:	3.14	15.7%	4	5	3	4	5	6	
(D4) Defensible Information Sources:	3.23	14.1%	3	6	4	2	5	1	

Table 3 (Cont.)

N = 78	"5-Point Scale" Mean Score Rating of Practice	Adjusted Percentage Who Rated Practice as "Fair-Poor"	Freq. of Reasons Given for "Fair-Poor"						
			1	2	3	4	5	6	
ACCURACY: (cont.)									
(D5) Valid Measurement:	2.65	48.7%	8	18	13	9	23	2	
(D6) Reliable Measurement:	2.74	40.0%	9	15	13	9	20	0	
(D7) Systematic Data Control:	2.93	30.4%	2	11	5	3	15	0	
(D8) Analysis of Quantitative Information:	2.92	30.6%	2	11	7	7	16	2	
(D9) Analysis of Qualitative Information:	2.86	33.8%	3	13	8	11	20	1	
(D10) Justified Conclusions:	2.79	40.3%	5	7	5	8	22	1	
(D11) Objective Reporting:	2.71	44.6	21	1	4	14	12	1	

Grouping of the Standards into Four Levels of Quality of Current Practice

	RANK	ITEM	(CATEGORY)	%RATING "FAIR-POOR"	$\bar{X}$ SCORE RATING
A. Satisfactory Practice Percentage "Fair-Poor": 14% : $\bar{X}$: 3.29	1. (C5)	- Rights of Human Subjects	(Propriety) :	4.3%	- 3.64
	2. (C4)	- Public's Right to Know	(Propriety) :	10.1%	- 3.46
	3. (C6)	- Human Interactions	(Propriety) :	14.7%	- 3.44
	4. (D4)	- Defensible Information Sources	(Accuracy) :	14.1%	- 3.23
	5. (D3)	- Described Purposes and Procedures	(Accuracy) :	15.7%	- 3.14
	6. (C8)	- Fiscal Responsibility	(Propriety) :	20.0%	- 3.15
	7. (D1)	- Object Identification	(Accuracy) :	19.1%	- 2.98
B. Fair Practice Percentage "Fair-Poor" : 31.1% : $\bar{X}$ 2.92	8. (C2)	- Conflict of Interest	(Propriety) :	29.5%	- 2.97
	9. (D7)	- Systematic Data Control	(Accuracy) :	30.4%	- 2.93
	10. (D8)	- Analysis of Quantitative Information	(Accuracy) :	30.6%	- 2.92
	11. (C3)	- Full and Frank Disclos.	(Propriety) :	31.0%	- 2.97
	12. (C7)	- Balanced Reporting	(Propriety) :	31.9%	- 2.94
	13. (B1)	- Practical Procedures	(Feasibility) :	30.3%	- 2.82
	14. (D9)	- Analysis of Qualitative Information	(Accuracy) :	33.8%	- 2.86
C. Marginal Practice Percentage "Fair-Poor" : 43.2% : $\bar{X}$: 2.70	15. (D10)	- Justified Conclusions	(Accuracy) :	40.3%	- 2.79
	16. (D6)	- Reliable Measurement	(Accuracy) :	40.0%	- 2.74
	17. (D2)	- Context Analysis	(Accuracy) :	39.7%	- 2.71
	18. (B3)	- Cost Effectiveness	(Feasibility) :	41.6%	- 2.70
	19. (A5)	- Report Clarity	(Utility) :	42.5%	- 2.70
	20. (C1)	- Formal Obligation	(Propriety) :	41.0%	- 2.66
	21. (D11)	- Objective Reporting	(Accuracy) :	44.6%	- 2.71
	22. (A2)	- Evaluator Credibility	(Utility) :	46.6%	- 2.73
	23. (A6)	- Report Dissemination	(Utility) :	47.1%	- 2.63
	24. (D5)	- Valid Measurement	(Accuracy) :	48.7%	- 2.65

Table 4 (cont.)

RANK	ITEM	CATEGORY	::	%RATING "FAIR-POOR"	—	$\bar{X}$ SCORE RATING
D. Poor Practice Percentage "Fair-Poor" : 60.2% : $\bar{X}$: 2.36						
25.	(A1) - Audiance Identification	(Utility)	:	50.0%	-	2.67
26.	(A3) - Information Scope and Selection	(Utility)	:	55.6%	-	2.43
27.	(A4) - Valuational Interpretation	(Utility)	:	57.1%	-	2.39
28.	(B2) - Political Viability	(Feasibility)	:	60.7%	-	2.28
29.	(A7) - Report Timeliness	(Utility)	:	63.4%	-	2.30
30.	(A8) - Evaluation Impact	(Utility)	:	74.6%	-	2.09

Table 5

Rating of Current Practice on the Standards by Role and Function of Respondents

N = 78	ROLE: Public School	Gov't Agency	Higher Educ.	Pvt. Firm	P Value	FUNCTION: Eval. Pract.	Admin. Eval.	Prog. Mgr.	2 + Func.	P Val
UTILITY:										
(A1) Audience Identification:	2.86	2.47	3.0	2.28		2.53	3.30	2.50	2.64	
(A2) Evaluator Credibility:	3.06	2.47	2.86	2.08	.0117	2.77	3.10	2.30	2.68	
(A3) Information Scope and Selection:	2.47	2.44	2.57	2.23		2.33	2.90	2.20	2.60	
(A4) Valuational Interpretation:	2.60	1.93	2.57	2.23		2.40	2.10	2.20	2.60	
(A5) Report Clarity:	3.09	2.50	2.43	2.20	.0145	2.74	2.90	2.60	2.64	
(A6) Report Dissemination:	3.06	2.36	2.67	1.93	.0028	2.68	2.80	2.33	2.62	
(A7) Report Timeliness:	2.58	1.75	2.43	2.20	.0426	2.32	2.45	2.11	2.26	
(A8) Evaluation Impact:	2.31	1.57	2.50	1.93	.0308	2.04	2.36	1.71	2.14	
FEASIBILITY:										
(B1) Practical Procedures:	2.97	2.53	2.67	2.85		2.89	2.87	2.33	2.90	
(B2) Political Viability:	2.32	2.14	2.14	2.42		2.00	2.37	2.25	2.31	
(B3) Cost Effectiveness:	2.85	2.47	3.20	2.50		2.59	3.12	2.71	2.67	
PROPRIETY:	3.39	2.84	3.46	2.81		3.13	3.24	3.16	3.15	
(C1) Formal Obligation:	2.50	2.47	3.50	2.79		2.70	2.44	2.14	2.89	
(C2) Conflict of Interest:	3.06	3.0	3.0	2.71		3.0	2.80	2.78	3.11	

Table 5 (cont.)

	ROLE					FUNCTION				
	Public School	Gov't Agency	Higher Educ.	Pvt. Firm	P Value	Eval. Pract.	Admin. Eval.	Prog. Mgr.	2 + Func.	P Valu
PROPRIETY: (cont.)										
(C3) Full and Frank Disclosure:	3.29	2.69	3.0	2.53	.0219	2.86	3.09	3.22	2.95	
(C4) Public's Right to Know:	3.79	3.0	3.71	3.07	.0033	3.33	3.82	3.80	3.29	
(C5) Rights of Human Subjects:	3.88	3.26	4.0	3.36	.0199	3.75	3.73	3.67	3.45	
(C6) Human Interactions:	3.87	2.80	3.86	3.0	.0002	3.44	3.60	3.60	3.29	
(C7) Balanced Reporting:	3.32	2.69	3.0	2.36	.0144	2.93	3.20	2.89	2.95	
(C8) Fiscal Responsibility:	3.41	2.79	3.57	2.67	.0112	3.04	3.20	3.20	3.25	
ACCURACY:										
(D1) Object Identification:	3.03	3.08	3.0	2.77		3.04	2.89	2.60	2.17	
(D2) Context Analysis:	2.94	2.36	3.14	2.29	.0249	2.68	2.89	2.30	2.86	
(D3) Described Purposes and Procedures:	3.28	2.94	3.57	2.77		3.20	3.40	3.00	3.00	
(D4) Defensible Information Sources:	3.41	3.19	3.43	2.71	.0162	3.23	3.40	3.40	3.05	
(D5) Valid Measurement:	3.03	2.25	2.71	2.14	.0017	2.63	2.90	2.30	2.73	
(D6) Reliable Measurement:	3.09	2.29	2.57	2.43	.0126	2.62	2.73	2.33	3.10	
(D7) Systematic Data Control:	3.17	2.62	3.33	2.47	.0177	3.07	2.90	2.67	2.86	
(D8) Analysis of Quantitative Information:	3.15	2.81	3.0	2.47		3.03	3.09	2.80	2.73	

Table 5 (Cont.)

	Public School	Gov't Agency	Higher Educ.	Pvt. Firm	P Value	Eval. Prac.	Admin. Eval.	Prog. Mgr.	2 + Func.	P Valu
(D9) Analysis of Qualitative Information:	3.18	2.62	3.00	2.29	.0145	2.93	3.09	2.80	2.68	
(D10) Justified Conclusions:	3.21	2.43	2.60	2.27	.0018	2.79	2.18	2.78	2.58	
(D11) Objective Reporting:	2.97	2.37	2.83	2.46		2.81	2.67	2.67	2.60	

Appendix B

Questionnaire

AN INVITATION .....

for CAPE members to participate in an important research project.

The Joint Committee on Standards for Educational Evaluation, chaired by Dr. Daniel Stufflebeam, is a national task force which has recently proposed a set of standards for evaluating educational programs, projects, and materials.

The purpose of this survey is to obtain information regarding how CAPE members

1. Perceive the extent to which each standard is an ideal to be sought.
2. Perceive the degree to which each standard is currently performed.
3. Perceive the major reasons why the standards are currently not being met.

The results of this survey will be given to Dr. Stufflebeam, as well as to the CAPE Committee on Standards for Program Evaluation, chaired by Dr. Jim Cox, who will report the results to the membership.

Please take a few minutes to complete the survey and return it by September 15, 1980 to Warren Newman, Los Angeles County Education Center, Room 210, 9300 East Imperial Highway, Downey, California 90242.

Thank you,

David Marsh

David Marsh
University of Southern California
School of Education

Warren Newman

Warren Newman
President
California Association of
Program Evaluators

DIRECTIONS: Please check those descriptors which best describe your current situation.	FUNCTIONS as a			
	PRACTICING EVALUATOR	ADMINISTRATOR of EVALUATION UNIT	PROGRAM MANAGER	OTHER (Please Specify)
EMPLOYED BY:				
____ Local School District				
____ Public Community College or University				
____ Private or Parochial School, College, or University				
____ Private Company				
____ County Education Office				
____ State Department of Education				
____ Non-Education Governmental Agency				
____ Other (please specify) _____				

Directions:

UTILITY: The Utility Standards are intended to ensure that an evaluation will serve the practical information needs of given audiences. These standards are:

(A1) *Audience Identification:* Audiences involved in or affected by the evaluation should be identified, so that their needs can be served.

(A2) *Evaluator Credibility:* The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance.

(A3) *Information Scope and Selection:* Information collected should be of such scope and selected in such ways that pertinent questions about the object of the evaluation are addressed, and so that the information is responsive to the needs and interests of specified audiences.

(A4) *Valuational Interpretation:* The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the base value judgments are clear.

	DEGREE TO WHICH YOU BELIEVE THE STANDARD IS APPROPRIATE AND IMPORTANT AND SHOULD BE RETAINED			YOUR ASSESSMENT OF THE QUALITY OF CURRENT PRACTICE IN EVALUATING SCHOOL PROGRAMS AS YOU SEE IT APPLIED					FOR EACH PRACTICE YOU RATED LESS THAN GOOD, PLEASE RANK THE MAJOR REASON(S) FOR THE INADEQUACY						
	INCLUDE -- *Important as stated*	EXCLUDE	REVISE -- *Comment on Text*	EXCELLENT	VERY GOOD	GOOD	FAIR	POOR	BUREAUCRATIC PROCEDURAL CONSTRAINTS	TIME CONSTRAINTS	BUDGET CONSTRAINTS	PROGRAM PERSONNEL RESISTANCE	LEVEL OF EVALUATOR COMPETENCE	OTHER -- *Please specify*	
(A1)															
(A2)															
(A3)															
(A4)															